

# Analysis of Hypertext Isolation Techniques for Cross-site Scripting Prevention

Mike Ter Louw Prithvi Bisht V.N. Venkatakrishnan

**UIC** Department of  
UNIVERSITY OF ILLINOIS  
AT CHICAGO Computer Science  
COLLEGE OF ENGINEERING

# Outline

Motivation

Hypertext isolation

Design challenges

Conclusion

*“Cross-site scripting (XSS) attacks would always fail if the browser could know for absolute certain which scripts were legitimate and which were malicious.”*

Gervase Markham

Mozilla Foundation

[http://www.gerv.net/security/  
content-restrictions/](http://www.gerv.net/security/content-restrictions/)

Many web users invest a large degree of trust into sites they visit.

- People expect web applications to:
  - preserve data confidentiality
  - provide accurate information
  - protect authentication credentials
- Lack of assurances forces trust model upon us

Modern websites blend code and data from many sources into a single execution environment.

Content types commonly found in a web application:

- First-party user-generated content
- Second-party web services and information
- Third-party advertisements
- Peer-authored content

# Ex: Facebook.com aggregates content from many sources.

The screenshot shows a Facebook profile for Matthew David Colebourne. The page is divided into several sections:

- Navigation:** Profile, Friends, Networks, Inbox, home, account, privacy, logout.
- Search:** A search bar with a magnifying glass icon.
- Applications:** Photos, Groups, Events, Marketplace, qubox, coComment User Conversations, and more.
- Profile Picture:** A photo of Matthew David Colebourne.
- Profile Info:** Update my status..., Networks: London, Sex: Male, Interested In: Women, Relationship Status: It's Complicated, Birthday: June 26, Hometown: Wokingham, England, Religious Views: Atheist.
- Mini-Feed:** A section for updates, currently empty.
- Information:** A section for profile information, currently empty.
- Education and Work:** A section for education and work, currently empty.
- qubox:** A section for qubox content, currently empty.
- coComment User Conversations:** A section for coComment User Conversations, currently empty.
- Friends:** A section for friends, showing 25 friends and a "See All" link. Friends listed include Yasunari Goto, Alex Lusby Taylor, Liam Hastings, Charles Nouyrnt, Kristina Serafim, and Juan Luis Hortelano.
- Friends in Other Networks:** A section for friends in other networks, currently empty.
- Comments of Blackblade:** A comment section with a post titled "After the Launch: managing a business, managing a micro business, independent professionals, web entrepreneur community, corporate branding, small business consultant, small business coach, small business expert, small business marketing, small business help, set up a business blog, how to get more clients, small business blogging: 3 Signs of a Savvy Entrepreneur". The comment text reads: "» I agree completely ... particularly on the last point. The beauty of web-based businesses, in particular, is the quick and relatively easy access to a huge market space (nationally and internationally) that was previously much harder to reach. However, that means that any individual's ability to 'predict' the market is much reduced ... I'm sure, for example, that Skype had no idea they would be so successful in Poland prior to the event. So, I am always concerned about anyone who suggests that they know exactly what will work and what will not. I think that you need invention to produce

- Ads
- User
- Apps
- Peers
- Facebook

# Ex: Facebook.com aggregates content from many sources.

The screenshot shows a Facebook profile for Matthew David Colebourne. The page is divided into several sections:

- Navigation:** Profile, Friends, Networks, Inbox, home, account, privacy, logout.
- Search:** Search bar with a magnifying glass icon.
- Applications:** Photos, Groups, Events, Marketplace, qubox, coComment User Conversations, and more.
- Profile Picture:** A photo of Matthew David Colebourne.
- Profile Info:** Update my status..., Networks: London, Sex: Male, Interested In: Women, Relationship Status: It's Complicated, Birthday: June 26, Hometown: Wokingham, England, Religious Views: Atheist.
- Mini-Feed:** A section for recent updates.
- Information:** A section for profile information.
- Education and Work:** A section for educational and professional details.
- qubox:** A section for user-generated content.
- coComment User Conversations:** A section for user conversations.
- Friends:** A section for friends, showing 25 friends and a "See All" link. Profiles of Yasunari Goto, Alex Lusby Taylor, Liam Hastings, Charles Nouyrit, Kristina Serafim, and Juan Luis Hortelano are visible.
- Friends in Other Networks:** A section for friends from other social networks.
- Comments of Blackblade:** A comment section with a post titled "After the Launch: managing a business, managing a micro business, independent professionals, web entrepreneur community, corporate branding, small business consultant, small business coach, small business expert, small business marketing, small business help, set up a business blog, how to get more clients, small business blogging: 3 Signs of a Savvy Entrepreneur". The comment text reads: "» I agree completely ... particularly on the last point. The beauty of web-based businesses, in particular, is the quick and relatively easy access to a huge market space (nationally and internationally) that was previously much harder to reach. However, that means that any individual's ability to 'predict' the market is much reduced ... I'm sure, for example, that Skype had no idea they would be so successful in Poland prior to the event. So, I am always concerned about anyone who suggests that they know exactly what will work and what will not. I think that you need invention to produce

- Ads
- User
- Apps
- Peers
- Facebook

# Ex: Facebook.com aggregates content from many sources.

The screenshot shows a Facebook profile for Matthew David Colebourne. The page is divided into several sections:

- Profile Header:** Name "Matthew David Colebourne", "Update my status...", and basic info: Networks (London), Sex (Male), Interested In (Women), Relationship Status (It's Complicated), Birthday (June 26), Hometown (Wolverhampton, England), Religious Views (Atheist).
- Left Sidebar:** Search bar, Applications (Photos, Groups, Events, Marketplace, qubox, coComment User Conversations), and a "CHRISTMAS MARKETS BY TRAIN" advertisement with an "EXPRESS BOOKER" form.
- Main Content Area:** "I am online now.", Friends list (Yasunari Goto, Alex Lusby Taylor, Liam Hastings, Charles Nouyrit, Kristina Serafim, Juan Luis Hortelano), and a "Friends in Other Networks" section.
- Right Sidebar:** A "Comments of Blackblade" section containing a long text post about business management and market prediction.
- Bottom Navigation:** A list of content sources: Mini-Feed, Information, Education and Work, qubox, and coComment User Conversations, each with a close button (X).

- Ads
- User
- Apps
- Peers
- Facebook

Image from <http://blog.cocomment.com/wp/wp-content/facebook.gif>



# Ex: Facebook.com aggregates content from many sources.

The screenshot shows a Facebook profile for Matthew David Colebourne. The profile includes a cover photo, a profile picture, and a bio. The bio lists personal details such as location (London), sex (Male), relationship status (It's Complicated), birthday (June 26), hometown (Wolingham, England), and religious views (Atheist). The profile also features a mini-feed with posts from qubox and coComment User Conversations. The qubox post is a green box with white text, and the coComment User Conversations post is a green box with white text. The profile is also linked to various applications like qubox and coComment User Conversations.

facebook Profile edit Friends Networks Inbox home account privacy logout

Search

Applications edit

- Photos
- Groups
- Events
- Marketplace
- qubox
- coComment User Conversations
- more

CHRISTMAS MARKETS BY TRAIN

EXPRESS BOOKER

London

Select Destination

Depart: 01 Dec

Return: 01 Dec

Proceed

More booking options

Matthew David Colebourne

Update my status...

Networks: London

Sex: Male

Interested In: Women

Relationship Status: It's Complicated

Birthday: June 26

Hometown: Wolingham, England

Religious Views: Atheist

Edit My Profile

I am online now.

Mini-Feed

Information

Education and Work

qubox

coComment User Conversations

Friends

25 friends See All

Yasunari Goto Alex Lusby Taylor Liam Hastings

Charles Nouyrif Kristina Serafim Juan Luis Hortelano

Friends in Other Networks

Comments of Blackblade

After the Launch: managing a business, managing a micro business, independent professionals, web entrepreneur community, corporate branding, small business consultant, small business coach, small business expert, small business marketing, small business help, set up a business blog, how to get more clients, small business blogging: 3 Signs of a Savvy Entrepreneur

> I agree completely ... particularly on the last point.

The beauty of web-based businesses, in particular, is the quick and relatively easy access to a huge market space (nationally and internationally) that was previously much harder to reach.

However, that means that any individual's ability to 'predict' the market is much reduced ... I'm sure, for example, that Skype had no idea they would be so successful in Poland prior to the event.

So, I am always concerned about anyone who suggests that they know exactly what will work and what will not. I think that you need invention to produce

- Ads
- User
- Apps
- Peers
- Facebook

# Ex: Facebook.com aggregates content from many sources.

The screenshot shows a Facebook profile for Matthew David Colebourne. The profile includes a cover photo, a profile picture, and a bio. The bio lists various details: Networks (London), Sex (Male), Interested In (Women), Relationship Status (It's Complicated), Birthday (June 26), Hometown (Wokingham, England), and Religious Views (Atheist). The profile also features a Mini-Feed, Information, and Education and Work sections. The Mini-Feed includes a post from qubox and a post from coComment User Conversations. The Information section lists Friends (25 friends) and Friends in Other Networks. The Education and Work section lists qubox and coComment User Conversations. The coComment User Conversations section shows a comment from Blackblade discussing business management and marketing.

facebook

Profile edit Friends Networks Inbox home account privacy logout

Search

Applications edit

- Photos
- Groups
- Events
- Marketplace
- qubox
- coComment User Conversations
- more

CHRISTMAS MARKETS BY TRAIN

EXPRESS BOOKER

London

Select Destination

Depart: 01 Dec

Return: 01 Dec

Proceed

More booking options

Matthew David Colebourne

Update my status...

Networks: London

Sex: Male

Interested In: Women

Relationship Status: It's Complicated

Birthday: June 26

Hometown: Wokingham, England

Religious Views: Atheist

Edit My Profile

I am online now.

Mini-Feed

Information

Education and Work

- qubox
- coComment User Conversations

▼ Friends

25 friends See All

Yasunari Goto Alex Lusby Taylor Liam Hastings

Charles Nouyrnt Kristina Serafim Juan Luis Hortalano

▼ Friends in Other Networks

Comments of Blackblade

After the Launch managing a business, managing a micro business, independent professionals, web entrepreneur community, corporate branding, small business consultant, small business coach, small business expert, small business marketing, small business help, set up a business blog, how to get more clients, small business blogging: 3 Signs of a Savvy Entrepreneur

I agree completely ... particularly on the last point.

The beauty of web-based businesses, in particular, is the quick and relatively easy access to a huge market space (nationally and internationally) that was previously much harder to reach.

However, that means that any individual's ability to 'predict' the market is much reduced ... I'm sure, for example, that Sk-type had no idea they would be so successful in Poland prior to the event.

So, I am always concerned about anyone who suggests that they know exactly

- Ads
- User
- Apps
- Peers
- Facebook

Image from <http://blog.cocomment.com/wp/wp-content/facebook.gif>

# Ex: Facebook.com aggregates content from many sources.

The screenshot shows a Facebook profile for Matthew David Colebourne. The page is divided into several sections:

- Header:** Profile edit, Friends, Networks, Inbox, home, account, privacy, logout.
- Search:** Search bar.
- Applications:** Photos, Groups, Events, Marketplace, qubox, coComment User Conversations.
- Profile Picture:** A blue-tinted photo of Matthew David Colebourne.
- Status:** Update my status...
- Personal Info:** Networks: London, Sex: Male, Interested In: Women, Relationship Status: It's Complicated, Birthday: June 26, Hometown: Wolverhampton, England, Religious Views: Atheist.
- Mini-Feed:** Information, Education and Work, qubox, coComment User Conversations.
- Friends:** 25 friends, See All. A grid of friend avatars including Yasunari Goto, Alex Luby Taylor, Liam Hastings, Charles Nouynt, Kristina Serafin, and Juan Luis Hortalano.
- Comments of Blackblade:** A comment discussing business management and web-based businesses.

- Ads
- User
- Apps
- Peers
- Facebook

Image from <http://blog.cocomment.com/wp/wp-content/facebook.gif>

# Web application authors endeavor to confine code and data into protection domains.

Important for web app to enforce these constraints:

- Limit capabilities within a protection domain
- Limit inter-domain data flows

# Cross-site scripting is a general class of attack to breach domain protection measures.

Abstract view of cross-site scripting (XSS):

1. Attack code input to web application
  - Type 0 DOM level zero reflection
  - Type 1 Request reflection
  - Type 2 Stored/persistent
2. Web app does not sufficiently filter or sanitize input
3. Attack succeeds
  - domain protections breached
  - **trust violated**

# Detection and suppression of malicious web content are challenging tasks for a web app.

- Standard protection measures (i.e., *same-origin policy*) too crude to be useful
  - Web apps must “roll own” fine-grained security policy enforcement mechanism
- Content parsing performed inconsistently across browsers
  - Identification of potentially harmful script code is hard
  - No robust way to distinguish active “code” from passive “data” [Hansen, 2008]

# Web apps are better positioned to define policy rules (rather than enforce them).

- App developers have better knowledge of:
  - origins of all emitted content
  - capabilities that are (in)appropriate for outsourced content
  - (un)desirable interactions between protection domains

# A key insight is made by Jim, Swamy and Hicks in BEEP.

## Observation ([Jim et al., 2007])

*To safely embed unknown, untrusted content:*

- *Web applications should define policy-based constraints.*
- *Web browsers should enforce these policies.*

Note: We already rely on browser to enforce same-origin policy.



# Content restrictions are well justified.

In summary,

- modern web applications integrate content from variety of sources
- level of trust varies by content source and use context
- web apps well suited to define protection domain policies
- browsers best suited to enforce policies

# Outline

Motivation

**Hypertext isolation**

Design challenges

Conclusion

*"In talking with the browser companies there seems to be more and more interest in content restrictions.*

*... The obvious answer [is] use an iframe to isolate it. That, unfortunately, has all sorts of user experience issues.*

*... So the best alternative is to create something that tells the browser, 'If you trust me, trust me to tell you to not trust me.'"*

Robert Hansen (a.k.a. "RSnake")

Author, XSS Cheat Sheet

[http://ha.ckers.org/blog/20070811/  
content-restrictions-a-call-for-input/](http://ha.ckers.org/blog/20070811/content-restrictions-a-call-for-input/)

# How might content restrictions work?

1. Web app breaks document down into logical regions:
  - contents of HTML element  
e.g., `<div>...</div>`
  - value of HTML element attribute  
e.g., `href="..."`
2. Web app declares policy-based constraints per region:
  - Inline  
e.g., `<div policy="...">...</div>`
  - Remote  
e.g., HTTP header targeting region
3. Browser associates policies with regions
4. Browser composes constraints for nested regions
  - Most restrictive constraint applies
5. Browser enforces composite constraints

# Policy-based constraints are weak without robust policy targeting.

Sometimes, intended  $\neq$  actual policy enforcement region

- Spurious close tags
  
- Implied “omitted” close tags (i.e., malformed HTML)

# Policy-based constraints are weak without robust policy targeting.

Sometimes, intended  $\neq$  actual policy enforcement region

- Spurious close tags

Intended `<div policy="..."></div><script...></div>`

- Implied “omitted” close tags (i.e., malformed HTML)

# Policy-based constraints are weak without robust policy targeting.

Sometimes, intended  $\neq$  actual policy enforcement region

- Spurious close tags

**Intended** `<div policy="..."></div><script...></div>`

**Actual** `<div policy="...">_</div><script...></div>`

- Implied “omitted” close tags (i.e., malformed HTML)

# Policy-based constraints are weak without robust policy targeting.

Sometimes, intended  $\neq$  actual policy enforcement region

- Spurious close tags

**Intended** `<div policy="..."></div><script...></div>`

**Actual** `<div policy="...">_</div><script...></div>`

- Implied “omitted” close tags (i.e., malformed HTML)

**Intended** `<table><div policy="...">`

`</table><script...></div></table>`



# Policy-based constraints are weak without robust policy targeting.

Sometimes, intended  $\neq$  actual policy enforcement region

- Spurious close tags

**Intended** `<div policy="..."></div><script...></div>`

**Actual** `<div policy="...">_</div><script...></div>`

- Implied “omitted” close tags (i.e., malformed HTML)

**Intended** `<table><div policy="...">  
</table><script...></div></table>`

**Actual** `<table><div policy="...">  
_</table><script...></div></table>`

# Content restrictions require isolation of hypertext.

## Observation

- *Effective content restriction requires accurate targeting of policy-based constraints to web content regions.*
- *To accurately target, policy declarations must robustly convey the targeted region's precise textual extent to the policy enforcement mechanism.*
  - *We term this need **hypertext isolation**.*

# Outline

Motivation

Hypertext isolation

Design challenges

Conclusion

# Hypertext isolation mechanisms compromise on several properties.

Ideal hypertext isolation mechanism. . .

1. increases utility of content restrictions
2. degrades well in today's browsers
3. has no regression from existing methods
4. maximizes usability

Six proposed techniques for hypertext isolation were analyzed.  
Each fell short in one or more of these areas.

# Evaluated techniques fell into six categories.

1. Document separation
  - `<iframe src="..."></iframe>`
2. Request separation
  - `<div src="..."></div>`
3. Response partitioning
  - MIME Multipart/Related (MHTML)
4. Element content encoding
  - `<div src="data:..."></div>`
5. Tag matching
  - `<div tag="unique">...</div tag="unique">`
6. Character range encoding
  - `<?isolate src="data:...">`

# Transitioning to a new feature requires legacy browser support.

Observed failure modes in non-supported browsers:

**poor** No content rendered

**graceful** At *least* trusted document regions are rendered

**safe** At *most* trusted document regions are rendered

**best** Trusted fallback content rendered in place of untrusted content

# Ideally, hypertext isolation should not make matters worse.

Some evaluated techniques had drawbacks over existing methods.

- Additional rendering delays  
e.g., appending untrusted content
- Additional, intensive HTTP request operations  
e.g., `<div src="...">`
- Unreadable (to humans) hypertext  
e.g., base64 encoding

# Hypertext isolation should enable usable content restrictions.

Some interesting applications benefit from the ability to:

1. isolate any type of document region
  - HTML element contents
  - element attribute values
  - JavaScript tokens
2. use same syntax in all contexts (i.e., context-free)
  - Retrofitting existing web apps



Although necessary, hypertext isolation is hard to get right.

In summary,

- Hypertext isolation is required for policy-based capability restriction of web content
- Many outstanding proposals provide isolation
- Isolation techniques differ on key design compromises

# Outline

Motivation



Hypertext isolation

Design challenges

Conclusion

To conclude,

- It is vitally important to the security of web apps that hypertext isolation be standardized and universally supported.
- Careful compromises should be made to obtain a sound framework for web content restrictions.
- This will help web apps continue to evolve while minding security.

-  Hansen, R. (2008).  
XSS cheat sheet.  
<http://ha.ckers.org/xss.html>.  
Retrieved on May 22, 2008.
-  Jim, T., Swamy, N., and Hicks, M. (2007).  
Defeating script injection attacks with browser-enforced  
embedded policies.  
*In 16th International World Wide Web Conference, Banff, AB,  
Canada.*

Thanks for your attention!

Questions?

# HTML standards community takes action!

Ian Hickson, editor for Web Hypertext Application Technology Working Group (WHATWG), yesterday added to HTML 5 proposed standard:

1. sandbox attribute of `<iframe>` element
  - Allows specification of (default-deny) policies:
    - `allow-same-origin`
    - `allow-forms`
    - `allow-scripts`
  - `<iframe>` can not open modal dialogs or alerts
  - all plugins disabled within `<iframe>`
  - navigation restrictions
2. seamless attribute of `<iframe>` element
  - Layout of `<iframe>` flows seamlessly into surrounding document (similar to `<div>`)
  - CSS style rules cascade into `<iframe>`

[www.whatwg.org/specs/web-apps/current-work/#sandbox](http://www.whatwg.org/specs/web-apps/current-work/#sandbox)